

**DISEÑO DE UN MODELO ESTADÍSTICO DE PRONÓSTICO DE LA ACTIVIDAD
COMERCIAL DE BANDA ANCHA PARA UNA EMPRESA DE
TELECOMUNICACIONES DE LA CIUDAD DE BOGOTA**

David Alejandro Jiménez Viveros
Oscar Mauricio Cabrera Cuervo

Fundación Universitaria Los Libertadores
Programa de Estadística Aplicada
Bogotá D.C.
2017

**DISEÑO DE UN MODELO ESTADÍSTICO DE PRONÓSTICO DE LA ACTIVIDAD
COMERCIAL DE BANDA ANCHA PARA UNA EMPRESA DE
TELECOMUNICACIONES DE LA CIUDAD DE BOGOTÁ**

David Alejandro Jiménez Viveros
Oscar Mauricio Cabrera Cuervo

Trabajo de grado para optar por el título de Especialista en estadística aplicada

Fundación Universitaria Los Libertadores
Programa de Estadística Aplicada
Bogotá D.C.
2017

Nota de aceptación

El trabajo de grado titulado, DISEÑO DE UN MODELO ESTADÍSTICO DE PRONÓSTICO DE LA ACTIVIDAD COMERCIAL DE BANDA ANCHA PARA UNA EMPRESA DE TELECOMUNICACIONES DE LA CIUDAD DE BOGOTÁ realizado por los estudiantes DAVID ALEJANDRO JIMÉNEZ VIVEROS y OSCAR MAURICIO CABRERA CUERVO, cumple con los requisitos exigidos por la FUNDACIÓN UNIVERSITARIA LOS LIBERTADORES para optar al título de ESPECIALISTA EN ESTADÍSTICA APLICADA.

Firma del presidente del jurado

Firma del jurado

Firma del Jurado

Bogotá D.C., 2 de diciembre de 2017

Las directivas de la Fundación Universitaria los libertadores, los jurados calificadores y el cuerpo docente no son responsables por los criterios e ideas expuestas en el presente documento. Estos corresponden únicamente a los autores.

CONTENIDO

	pág.
1. CONTENIDO	5
2. INTRODUCCIÓN	9
3. OBJETIVOS	11
Objetivo general.	11
Objetivos específicos.	11
4. MARCO TEÓRICO	12
2.1. RTC Red telefónica conmutada	13
2.2. banda ancha	14
2.3. Series de tiempo	15
2.4. Modelo Box-Jenkis	16
2.5. Modelos Autoregresivos Vectoriales (VAR)	18
2.6. Pruebas En Modelos.	19
5. MARCO METODOLÓGICO	20
6. ANÁLISIS Y RESULTADOS	33
7. CONCLUSIONES Y RECOMENDACIONES	36
8. REFERENCIAS	37

LISTA DE TABLAS

	pág.
TABLA 1: ANÁLISIS DESCRIPTIVO DE LOS DATOS	20
TABLA 2: MATRIZ DE CORRELACION DE LAS VARIABLES ORIGINALES	21
TABLA 3: TEST DICKEY-FÜLLER AUMENTADO SERIE ORIGINAL	23
TABLA 4: BIC DE LOS MODELOS ESTIMADOS	24
TABLA 5: MODELOS SUGERIDOS POR VARSELECT	28
TABLA 6: TEST DE CAUSALIDAD	31
TABLA 7: PRUEBAS DE CORRELACIÓN SERIAL DE LOS RESIDUALES DEL MODELO VAR	31

LISTA DE FIGURAS

	pág.
FIGURA 1: TOPOLOGÍA DE UNA RTC	13
FIGURA 2: PROCESO BOX JENKIS	17
FIGURA 3: DIAGRAMA DE CAJAS VARIABLES ORIGINALES	21
FIGURA 4: DIAGRAMA DE CAJAS VARIABLES ESTANDARIZADAS	21
FIGURA 5: SERIE ALTAS	22
FIGURA 6: ACP Y PACF ALTAS	22
FIGURA 7: SERIE ALTAS TRANSFORMADA	23
FIGURA 8: PRUEBAS BOX LJUNG Y JARQUE BERA	25
FIGURA 9: DATOS REALES VS DATOS PRONOSTICADOS	26
FIGURA 10: SERIE VARIABLES SELECCIONADAS	26
FIGURA 11: ACF Y PACF DE VARIABLES PRINCIPALES	27
FIGURA 12: ACF Y PACF DE LOS RESIDUALES DE LA VARIABLE ALTAS PARA EL MODELO VAR(2)	29
FIGURA 13: ACF Y PACF DE LOS RESIDUALES DE LA VARIABLE VENDEDORES PARA EL MODELO VAR(2)	29
FIGURA 14: ACF Y PACF DE LOS RESIDUALES DE LA VARIABLE INGRESOS PARA EL MODELO VAR(2)	30
FIGURA 15: ACF Y PACF DE LOS RESIDUALES DE LA VARIABLE GASTOS PARA EL MODELO VAR (2)	30
FIGURA 16: PREDICCIÓN MODELO VAR(2)	32
FIGURA 17: COMPARACIÓN DEL MODELO SARIMA(010)(110)12 CON MODELO VAR(2)	33
FIGURA 18: PREDICCIÓN DE ALTAS AÑO 2018	34
FIGURA 19: ESTIMACIÓN DE INVERSIÓN	34
FIGURA 20: PROYECCIÓN VENDEDORES, INGRESOS Y GASTOS	35

LISTA DE ECUACIONES

	pág.
ECUACIÓN 1: REPRESENTACIÓN MATEMÁTICA DE UNA SERIE DE TIEMPO UNIVARIADA.	15
ECUACIÓN 2: REPRESENTACIÓN MATEMÁTICA DE UNA SERIE DE TIEMPO MULTIVARIADAS	15
ECUACIÓN 3: MODELO SARIMA	18
ECUACIÓN 4: MODELO VAR	18
ECUACIÓN 5: SARIMA(010)(110)12	33
ECUACIÓN 6: VAR2	35

GLOSARIO

Altas: Ventas del servicio de banda ancha

Vacancia: Porcentaje de red libre

Internet: Redes informáticas interconectadas que utilizan los protocolos TCP/IP que permite compartir información, en cualquier formato sea texto, audio, imágenes o videos.

TCP/IP: Siglas de protocolo de transmisión / protocolo de internet que sirve para comunicarse entre los diferentes sistemas operativos, Ejemplo: Windows, Apple, Linux, entre otros.

Banda Ancha: Red dedicada que permite la conexión a internet para la transmisión o recepción de datos.

VAR: Modelo auto regresivo de vectores.

Kbps: Medida de transmisión de datos en Kilobyte, 1Kbps son 1024 Bytes.

Mbps: Medida de transmisión de datos en Megabytes 1Mbps son 1024 Kilobytes.

RDSI: Red digital de servicios integrados.

xDSL: Medio de conexión a la internet a través de líneas telefónicas de cobre.

Resumen

La comercialización de productos de banda ancha es un mercado muy competitivo el cual depende de muchos factores, uno de los principales son las capacidades de infraestructura de red. La identificación de los periodos de inversión de red en algunos casos no es la adecuada al no utilizar modelos estadísticos, por tal motivo se plantea el diseño de un modelo estadístico de pronóstico basado en modelos univariados SARIMA y multivariados VAR comparando su eficiencia en la proyección de altas y pudiendo identificar el momento de inversión respecto al inventario de red.

1. INTRODUCCIÓN

Planteamiento del problema

La falta de análisis de la información en las empresas de telecomunicaciones genera una ineficiencia en las labores comerciales generando pérdida de tiempo, dinero y el no aprovechamiento de los recursos físicos. La inversión que se realiza en los recursos de red para poder prestar servicios de telecomunicaciones es demasiado costosa, en el despliegue y su mantenimiento; dicha inversión debe ser retornada por medio de una actividad comercial la cual se enfoca, en información histórica, disponibilidad de oferta y vacancia en los circuitos de red; la ausencia de un modelo estadístico en las actividades de comercialización de los servicios de banda ancha conlleva a tomar decisiones erradas que comprometen el capital de las empresas de telecomunicaciones.

Formulación

Las actividades comerciales de los productos y servicios marcan diferencias en los ingresos de todas las empresas, el poder enfocar las ventas con eficiencia es el principal objetivo que se busca en los negocios que tienen involucrados recursos físicos. El mercado de las telecomunicaciones es un claro ejemplo de competitividad comercial en donde el avance tecnológico y la demanda de servicios sobre las redes existentes, obligan a las empresas a tener estrategias de mercado, garantizando el aprovechamiento de los recursos. Los servicios de acceso a internet sobre redes fijas o móviles tienen la mayor demanda en el sector de las telecomunicaciones, las estrategias de comercialización de estos servicios se enfocan en el medio de acceso y velocidad ofrecida, dependiendo de la capacidad de los recursos de red y creando la necesidad de una buena administración, de la información con planes de comercialización enfocados en los recursos de red libres sobre las zonas con demanda.

En Colombia la apertura de los mercados se dio con la privatización de las principales empresas de telecomunicaciones, abriendo las puertas a multinacionales para la inversión en nuevas tecnologías generando un mercado con gran demanda y masificando los servicios de internet conmutado seguido de la banda ancha, “la competencia convierte los esfuerzos por el lucro individual en beneficios para toda la sociedad” (Sarmiento, 2017), esto favorece el crecimiento de este servicios con base a la competencia. En la actualidad toda la inversión y expansión realizada por las empresas de telecomunicaciones se basa en infinidad de variables como la demanda, recursos de red, presupuestos asignados, competencia, situación socioeconómica, etc. Una de las plazas con mayor demanda

del servicio de banda ancha es Bogotá registrando “penetración de internet fijo con 21,6 por ciento” (Tecnóesfera, 2017), donde la densidad de sus clientes y la saturación de las redes de todos los operadores complican las actividades de comercialización, donde lleva a cuestionar si las actividades comerciales para las ventas de servicios de banda ancha se pueden predecir a través de un modelo estadístico, donde se pueda tener el comportamiento de las ventas, anticipándose a la competencia y a la saturación de recursos para realizar inversión en las redes.

Justificación.

Las empresas de telecomunicaciones en la ciudad de Bogotá comercializan el servicio de banda ancha con base en las demandas del sector y las capacidades de disponibilidad de la red, los factores que pueden afectar las ventas tienen un análisis pobre, basado en promedios y no cuenta con análisis más completos que permitan realizar un despliegue ordenado de las ventas. Los presupuestos son asignados con base en estudios de mercado contemporáneos, sin tener en cuenta el comportamiento histórico de muchas variables que pueden afectar las ventas, no hay un análisis de la relación entre los diferentes factores que involucran las actividades de comercialización.

En este proyecto se pretende diseñar un modelo estadístico de pronóstico en la ciudad de Bogotá, basado en la información histórica de la compañía e identificando las principales variables que nos permitan estimar de manera anticipada el comportamiento de las actividades de comercialización para el producto de banda ancha.

2. OBJETIVOS

OBJETIVO GENERAL.

Diseñar un modelo estadístico que permita estimar de manera anticipada el comportamiento de las altas en la actividad comercial, para la planeación de inversión de red.

OBJETIVOS ESPECÍFICOS.

- Identificar las fuentes de información y variables que involucran las actividades comerciales de banda ancha.
- Estimar y comparar los modelos SARIMA y VAR determinando el más adecuado para las actividades de altas.
- Generar el pronóstico de altas, identificando el mes en que se debe realizar la inversión en infraestructura de red en el año 2018.

3. MARCO TEÓRICO

La necesidad de intercambiar información ha sido el factor primordial para el desarrollo de medios tecnológicos en las telecomunicaciones, pasando desde el correo postal seguido de las comunicaciones telefónicas hasta la revolución de Internet, con el cual tenemos acceso a toda la información que el hombre ha digitalizado. Internet nació en los años sesenta en los Estados Unidos por desarrollos militares a consecuencia de las diferencias entre las potencias del momento (Estados Unidos y la Unión Soviética) su enfoque era tener acceso a la información desde cualquier parte del mundo; en el año 1971 se efectuaron la primeras conexiones de computadoras a la red ARPANET realizando el primer envío de un correo electrónico, desde ese momento el crecimiento de las conexiones generó la necesidad de diseñar un nuevo protocolo de comunicación llamado TCP/IP que se convirtió en el estándar de las comunicaciones informáticas, permitiendo el intercambio de información a nivel mundial abriendo las puertas a todas las actividades comerciales del mundo e impulsando las economías haciendo más fácil la comercialización de cualquier producto entre países.

Internet llevó a las empresas telefónicas existentes a ofrecer accesos por medio de redes de cobre, ofreciendo un servicio de internet conmutado a través de un modem con una marcación al ISP (proveedor de servicios de internet) quien realizaba una conexión analógica virtual a internet en velocidades de 14kbps, 28kbps, 33kbps y 56kbps; estas conexiones eran muy inestables, lentas y no se podía utilizar el servicio telefónico mientras se realizaba la conexión a internet, a raíz del crecimiento de los servicios y contenidos de internet las conexiones requerían de mayor velocidad, para esto las empresas de telefonía ofrecieron un producto llamado RDSI con una conexión digital soportando velocidades de 128kbps hasta 2Mbps, estos servicios permitían usar los servicios de red mientras existía una llamada telefónica, eran bastante costosos y por lo general lo contrataban clientes comerciales.

Las empresas telefónicas se convirtieron en lo que actualmente se conoce como empresas de telecomunicaciones y mejoraron los servicios de conexión a internet optimizando las conexiones, apoyándose en las tecnologías xDSL que trabajan sobre las redes de cobre existentes pudiendo llegar hasta velocidades de 100Mbps mejorando la experiencia de los clientes y garantizando un servicio estable. La evolución de estas tecnologías atendiendo la demanda de Internet a altas velocidades, teleconferencias, televisión por demanda, video comunicaciones, entre muchas otras aplicaciones da inicio al servicio de banda ancha.

2.1 RTC RED TELEFÓNICA CONMUTADA

La red telefónica conmutada o RTC es la estructura básica para prestar servicios de telecomunicaciones, esta se compone de elementos de transmisión y conmutación que están interconectados entre sí por redes de cobre y fibra óptica permitiendo interconectar equipos de los suscriptores como teléfonos, computadores, televisores, etc. Los elementos de estas redes son:

- Central telefónica (distribuidor general).
- Canalizaciones. Cámaras. Cables Primarios / de Red Directa.
- Puntos de Distribución (ARMARIOS), bloques primarios y bloques secundarios.
- Cables Secundarios. Puntos de Subrepartición (cajas de dispersión, strip).

Todos estos elementos tienen diversas tareas dentro de las RTCs y pueden ser activos como las centrales telefónicas que por medio de software realizan la conmutación y transmisión de todas las peticiones de comunicación incluyendo acceso a Internet, o pasivos en el caso de los cables de cobre los cuales están tendidos en las ciudades utilizando infraestructuras públicas hasta llegar a los puntos de dispersión ubicados en los postes telefónicos donde por medio de un par de cobre se conecta a los suscriptores, estos elementos los podemos observar en la Figura 1. Es de aclarar que todas estas redes tienen una capacidad y por cada cliente que se conecta a esta, ocupa un recurso dedicado, y se resalta la importancia de tener un adecuado administrador de recursos de red.

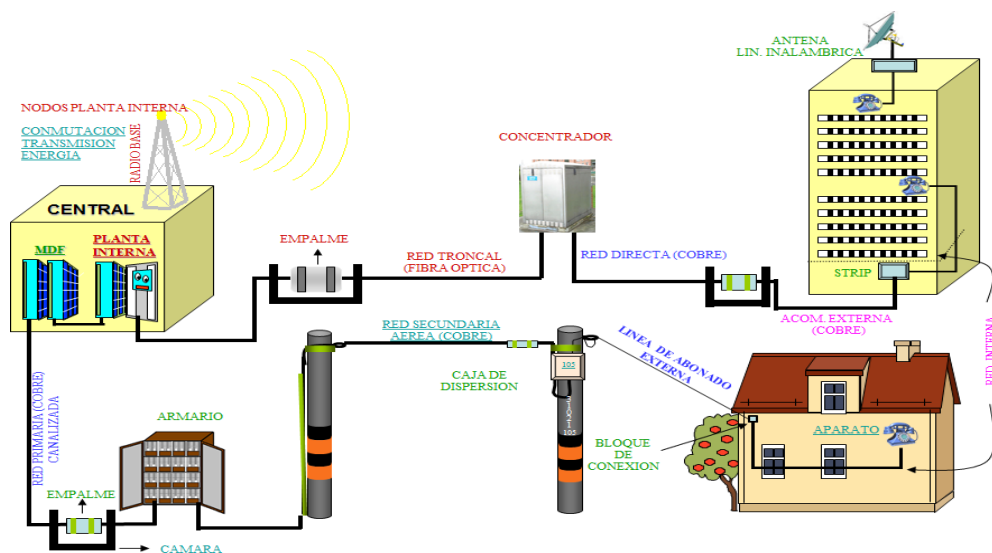


Figura 1: Topología de una RTC

Dentro de las topologías de red de cobre para una RTC podemos encontrar una clasificación por su función:

- Redes troncales (conectan centrales).
- Red de abonado (conectan a los clientes).

Respecto a su instalación en:

- Red aérea (cables de cobre tendidos en postes de concreto).
- Red canalizada (cables de cobre subterráneos a través de ductos y cámaras).

La red de abonado se clasifica en:

- Red directa (sin armario de dispersión)
- Red flexible (con armario de dispersión, y consta de dos sub partes, red primaria y red secundaria).

En la actualidad las redes de cobre que empezaron con solo telefonía son la base de los servicios de banda ancha en donde la comercialización de este servicio es un reto para las empresas del sector y depende de las capacidades y proyección de sus redes, para fines prácticos las cantidades de los inventarios de red con posibilidad para las instalaciones de un servicios de banda ancha se les conoce como circuitos y este es el inventario de red.

2.2. BANDA ANCHA

La banda ancha o acceso a Internet de alta velocidad permite a los usuarios acceder a Internet y a los servicios relacionados a velocidades significativamente más rápidas que las disponibles mediante los servicios conmutados. Las velocidades de banda ancha varían desde 1Mbps hasta 50Mbps, dependiendo de la tecnología y del nivel de servicio contratado. Los servicios de banda ancha para consumidores residenciales proveen típicamente velocidades más rápidas de descarga de datos que las velocidades de carga de datos. En Colombia, los cinco proveedores que presentaron el mayor número de suscriptores con acceso de Internet fijo fueron: Telmex 2'237.650, UNE EPM 1'238.613, Colombia Telecomunicaciones 974.311, Empresa de Telecomunicaciones de Bogotá 662.940 y Edatel S.A. E.S.P 197.841 (Tecnóesfera, 2017).

La banda ancha realiza una comunicación digital y puede realizarse por diversos medios como, redes de cobre, fibra óptica e inalámbrica. La tecnología de acceso más usada por operadores en Colombia es el cable con 3'140.799 suscriptores con un 52% de participación, el departamento con mayor penetración de banda ancha fija es Bogotá con un 21.8% (Tecnóesfera, 2017). La meta para el 2019 es llegar a un estándar de banda ancha situado en 25 Mbits de velocidad.

Podemos asegurar que la banda ancha es uno los principales servicios de las empresas de telecomunicaciones, pero este depende de una infraestructura limitada que es punto clave en la comercialización de este producto, creando la necesidad de generar planes estratégicos para la venta, mantenimiento del servicio y estimación de inversión de red.

2.3. SERIES DE TIEMPO

Una serie temporal (o simplemente una serie) es una secuencia de N observaciones (datos) ordenadas y equidistantes cronológicamente sobre una característica (serie univariante o escalar) o sobre varias características (serie multivariante o vectorial) de una unidad observable en diferentes momento (Mauricio, 2007).

$$\{\gamma_1, \gamma_2, \dots, \gamma_T\}; \{\gamma_t\}_{t=1}^T, \text{ donde } \gamma_t \text{ es la observación } n^{\circ} t \\ (1 \leq t \leq T) \text{ y } T \text{ es el número de observaciones.}$$

Ecuación 1: Representación matemática de una serie de tiempo univariada.

$$\{\gamma_1, \gamma_2, \dots, \gamma_T\}; \{\gamma_t\}_{t=1}^T, \text{ donde } \gamma_t \equiv [\gamma_{t1}, \gamma_{t2}, \dots, \gamma_{tM}]' (M \geq 2) \text{ es la observación } n^{\circ} t \\ (1 \leq t \leq T) \text{ y } T \text{ es el número de observaciones.}$$

Ecuación 2: Representación matemática de una serie de tiempo multivariadas

Las series de tiempo pueden ser enfocadas en descripción, explicación, predicción y control.

- Descripción: la gráfica de la series original nos puede mostrar sus características respecto a su comportamiento en el tiempo, aquí podremos observar comportamientos regulares en el tiempo o tendencias de crecimiento o decrecimiento, que nos pueden indicar o describir el comportamiento de los eventos observados.
- Explicación: al analizar más de una variable podremos observar la variación de una serie de tiempo respecto a la otra y poder entender cómo se afectan utilizando modelos de regresión múltiple.

- Predicción: al tener registros de eventos en el tiempo y analizarlos podríamos realizar una predicción o realizar un pronóstico acerca de su comportamiento en el futuro, esto nos puede ayudar a saber con anterioridad cuando podríamos tener un evento especial o atípico.
- Control: en los procesos industriales las series de tiempo se pueden encaminar en la toma de información de procesos, comparándolas con las caracterizas de calidad para tomar decisiones respecto a los productos o maquinaria utilizados.

Para analizar las series de tiempo se debe tener en cuenta si hay tendencia, variación estacional, ciclos de cambio y variaciones atípicas o irregulares, estos nos permiten identificar los pasos a seguir en la asignación de un modelo. En algunas series no podemos identificar a simple vista dichos comportamientos, para estos casos existen pruebas que nos permiten identificarlos (Chatfield, 1995).

Las series originales pueden transformarse con logaritmos o raíces cuadradas cuando tengamos que estabilizar la varianza, convertirla en serie estacionaria o distribución normal.

2.4 MODELO BOX-JENKIS

El análisis de series de tiempo enfocadas en la predicción por parte de George Edward Pelham Box y Gwilym Jenkins, “marcó el comienzo de una nueva generación de herramientas de pronóstico. Popularmente conocida como metodología de Box-Jenkins (BJ), pero técnicamente conocida como metodología ARIMA” (Gujarati & Porter, 2010, pág. 774). Los modelos ARIMA o modelos autoregresivos integrados de media móvil tienen el objetivo de analizar las propiedades estocásticas de las series de tiempo, donde los datos históricos pueden explicar el comportamiento de algún fenómeno o proceso. Los modelos ARIMA se utilizan en series de tiempo univariadas (con una sola variable en el tiempo) pero se pueden extender a modelos multivariados.

El objetivo primordial de los modelos ARIMA es usar datos estadísticos para realizar predicciones a corto y mediano plazo, utilizando los pasos de la Figura 2, en estos modelos también podemos encontrar estacionalidad en periodos, para estos casos se denominan modelos SARIMA.

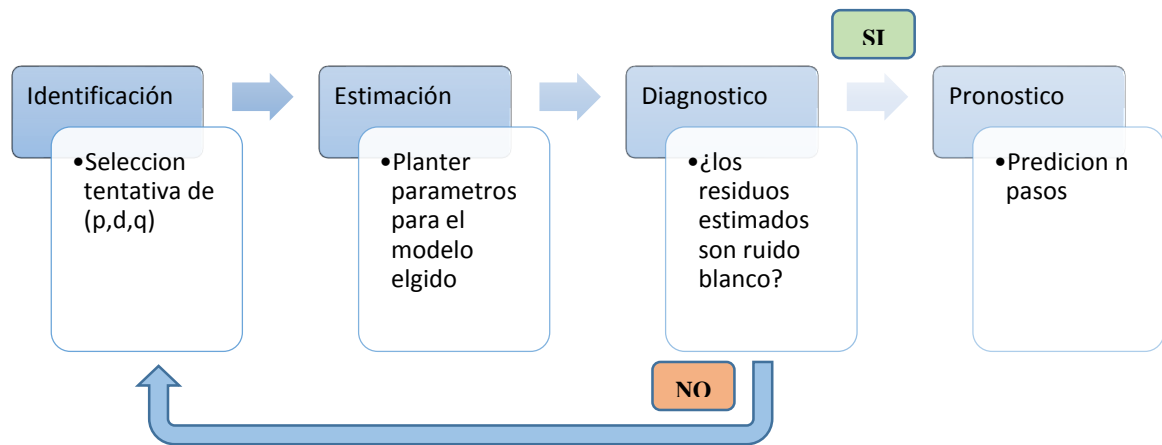


Figura 2: Proceso Box Jenkins

A continuación se describen los pasos del modelo Box Jenkins:

- **Identificación:** Se identificamos si debemos aplicar alguna transformación a los datos y los posibles valores de orden para el modelo, utilizando la autocorrelación simple y parcial.
- **Estimación:** Se estiman los parámetros de los términos autorregresivos y de medias móviles incluidos en el modelo. “Algunas veces, este cálculo se efectúa mediante mínimos cuadrados simples, pero otras hay que recurrir a métodos de estimación no lineal (en parámetros). Como esta labor se lleva a cabo ahora a través de rutinas en diversos paquetes estadísticos” (Gujarati & Porter, 2010, pág. 777).
- **Diagnóstico:** En esta etapa se procede a evaluar el modelo seleccionado verificando si se ajusta a los datos originales. Para validar si es el mejor modelo, debemos garantizar que los residuales estimados sean ruido blanco, de lo contrario debemos buscar otro modelo. Pueden existir varios modelos que se ajusten, “es por esto que el diseño de modelos ARIMA de Box-Jenkins es un arte más que una ciencia; se requiere gran habilidad para seleccionar el modelo ARIMA correcto” (Gujarati & Porter, 2010, pág. 777). También se evalúan por medios de pruebas de hipótesis, IAC y BIC. Cuando no se logre que los residuales sean ruido blanco se debe regresar al paso de estimación hasta encontrar el modelo adecuado.
- **Etapas de pronóstico:** En este paso utilizamos el modelo que más se ajustó en el diagnóstico cumpliendo con los requisitos de las pruebas y procedemos a realizar un pronóstico.

En el desarrollo de este proyecto se utiliza el modelo univariado SARIMA, descrito a continuación.

Modelo SARIMA

El modelo SARIMA es utilizado para las series de tiempo univariadas donde en un previo análisis de estimación basado en Box Jenkins se evidencia una estacionalidad en la serie. Para este tipo de modelos se tiene que modelar la parte regular o no estacional y la parte estacional, convirtiéndose en un modelo producto de dos ARIMA(p,d,q) x SARIMA(P,D,Q). Las series que se trabajan pueden ser estacionarias en su parte regular y también pueden tener tendencia en su parte estacional, por lo tanto pueden requerir diferenciación en cada una de ellas (Chatfield, 1995). En la Ecuación 7 vemos el modelo SARIMA.

$$\begin{aligned} \phi_p(B)\Phi_P(B^S)W_t &= \theta_q(B)\Theta_Q(B^S)e_t \\ \text{Donde} \\ \phi_p, \Phi_P, \theta_q, \Theta_Q &\text{ son polinomios de orden } p, P, q, Q \\ W_t &= \nabla^d \nabla_s^D X_t \end{aligned}$$

Ecuación 3: Modelo SARIMA

Los modelos SARIMA siguen el proceso de Box Jenkins descrito en la figura 2, donde se siguen todos los pasos hasta encontrar el modelo que mejor se ajuste.

2.5 MODELOS AUTOREGRESIVOS VECTORIALES (VAR)

El análisis VAR ha evolucionado como un instrumento estándar en econometría para el análisis de series temporales multivariadas. Debido a que las pruebas estadísticas son muy utilizadas para determinar la interdependencia y las relaciones dinámicas entre variables, pronto se hizo evidente que la metodología podría enriquecerse incorporando información no estadística a priori (TSAY, 2013). Los modelos autoregresivos VAR se utilizan para caracterizar las interacciones simultáneas entre un grupo de variables. VAR es apropiado cuando existe evidencia de simultaneidad entre un grupo de variables, y que sus relaciones se transmiten a lo largo de un determinado número de períodos. Este modelo consiste en un grupo de variables K en el tiempo, donde:

$$\begin{aligned} y_t &= (y_{1t}, \dots, y_{1kt}, \dots, y_{Kt}) \text{ Para } K=1, \dots, K. \\ y_t &= A_1 y_{t-1} + \dots + A_p y_{t-p} + \varepsilon_t, \text{ donde } A \text{ es } (K \times K) \text{ matrices,} \\ &\text{Para } i = 1, \dots, p \\ y_t &= A_t y_{t-1} + v_t, \end{aligned}$$

Ecuación 4: Modelo VAR

2.6 PRUEBAS EN MODELOS.

Para poder determinar transformaciones de las series originales y confirmar si los modelos son apropiados para predicciones se pueden realizar las siguientes pruebas:

Prueba Dickey-Füller (DF): Esta consiste en determinar la existencia o no de raíces unitarias, determinado si la serie no es estacionaria, se realiza a través de una prueba T donde si rechazamos la hipótesis de que $\delta = 0$ la serie es estacionaria (Gujarati & Porter, 2010, pág. 755). Esta prueba también tiene una versión aumentada (ADF) en donde se utiliza un número negativo en la prueba de hipótesis, entre más negativo es el número más fuerte es el rechazo de la hipótesis nula de que existe una raíz unitaria para un cierto nivel de confianza.

Ljung-Box (LBQ): esta prueba permite comprobar si una serie de observaciones en un periodo de tiempo específico son aleatorias e independientes. Si las observaciones no son independientes, una observación puede estar correlacionada con otra observación k unidades de tiempo después, una relación que se denomina autocorrelación. La autocorrelación puede reducir la exactitud de un modelo predictivo basado en el tiempo, como la gráfica de series de tiempo, y conducir a una interpretación errónea de los datos. Donde por medo de una prueba chi cuadrado se evalúa la hipótesis nula corresponde a que los datos se distribuyen de forma independiente y la alternativa que los datos no se distribuyen de forma independiente.

Prueba de normalidad de Jarque-Bera (JB): Esta calcula primero la asimetría y la curtosis de los residuos de MCO, sigue la distribución chi cuadrada, con 2 grados de libertad. Si el valor p calculado del estadístico JB es lo bastante bajo en una aplicación, lo cual sucederá si el valor del estadístico difiere en gran medida de cero, se puede rechazar la hipótesis de que los residuos están normalmente distribuidos (Gujarati & Porter, 2010, pág. 132).

Prueba LM de correlación serial: el Test del multiplicador de Lagrange de correlación serial de Breusch–Godfrey identifica si hay dependencia serial dentro de un modelo propuesto, si la hubiese llevara a pronósticos incorrectos, la hipótesis nula indica que no hay correlación serial en los residuos, por lo tanto, el objetivo para que el modelo sea óptimo es no rechazarla.

4. MARCO METODOLÓGICO

Para abordar el proyecto de pronóstico de la actividad comercial de una empresa de telecomunicaciones en la ciudad de Bogotá se identifican los sistemas administradores de información de clientes y de red, en los cuales se guardan las cantidades de actividades de ventas y el detalle de red. Dentro de estos sistemas administraciones se seleccionaron las siguientes variables.

- **Altas:** cantidad de ventas efectivas del producto de banda ancha en el mes.
- **Vendedores:** Personal de ventas en el mes
- **Ingresos:** Ingreso en pesos colombianos del primer mes, producto de la venta de banda ancha.
- **Gastos:** Gasto asociados a las ventas del mes
- **Mix de velocidad:** Velocidad media de altas en Mbits del mes.
- **Mix de estrato:** Estrato medio de altas del mes

Con base en las variables que se identificaron se capturaron los datos con un periodo de 82 meses desde enero de 2011 hasta octubre de 2017.

Se realiza análisis exploratorio de las variables seleccionadas, encontrando las siguientes características:

	Altas (Qxm)	Vendedores (Qxm)	Ingresos (MCOPxm)	Gastos (MCOPxm)	MixVelocidad (Prom mes)	MixEstrato (Prom mes)
Min	1996	594	82,39	135,1	2,180	1,76
1Q	3488	742	112,88	173,5	2,982	2,33
Mediana	4030	823	125,83	202,9	4,485	2,5
Media	3985	816,9	125,28	215,4	5,029	2,443
3Q	4474	885,5	136,87	259,2	5,770	2,627
Max	6257	1033	167,42	325,5	11,05	2,8

Tabla 1: Análisis descriptivo de los datos

Apoyándonos en los diagramas de cajas de las figura 3 y 4 identificamos que los gastos tienen mayor varianza y existen datos atípicos para altas, mix de velocidad y mix de estratos, procedemos a mirar las correlaciones entre variables identificando las principales.

	Altas	Vendedores	Ingresos	Gastos	MixVelocidad	MixEstrato
Altas	1	0,7713594	0,7615487	0,5852704	0,5377845	0,2154579
Vendedores	0,7713594	1	0,68946	0,842229	0,5185981	0,1787679
Ingresos	0,7615487	0,68946	1	0,4773192	0,3405537	0,1496806
Gastos	0,5852704	0,842229	0,4773192	1	0,7771709	0,4230969
MixVelocidad	0,5377845	0,5185981	0,3405537	0,7771709	1	0,4953311
MixEstrato	0,2154579	0,1787679	0,1496806	0,4230969	0,4953311	1

Tabla 2: Matriz de correlación de las variables originales

Según la matriz de correlación las variables más significativas respecto a las altas son los vendedores, ingresos y gastos.

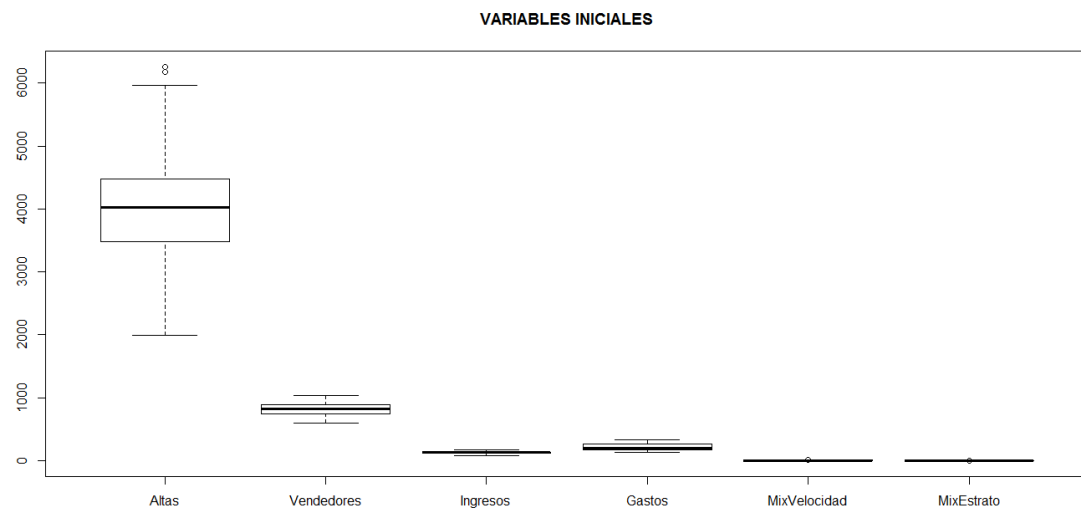


Figura 3: Diagrama de cajas variables originales

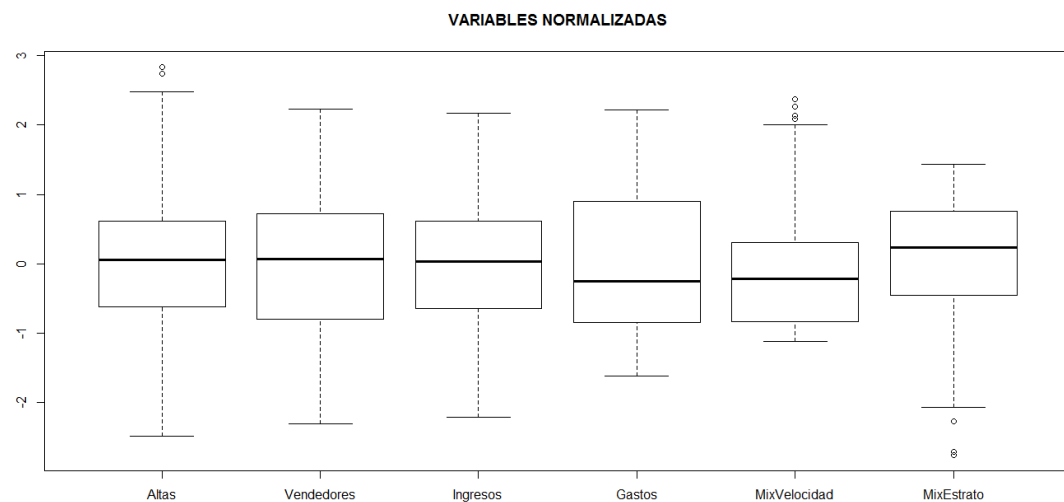


Figura 4: Diagrama de cajas variables estandarizadas

Identificando las variables que se van a incluir al modelo: Altas, Vendedores, Ingresos y Gastos, se procede a utilizar la metodología Box-Jenkins para desarrollar el modelo predictivo.

Modelo SARIMA

Identificación:

Se procede a realizar las gráficas de cada una de las variables en el tiempo para realizar la identificación del comportamiento de los datos y orden de los posibles (p,q) .

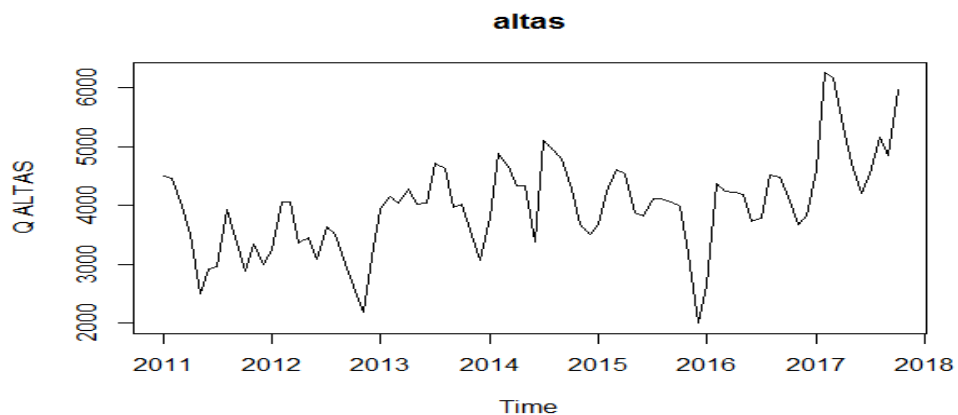


Figura 5: Serie altas

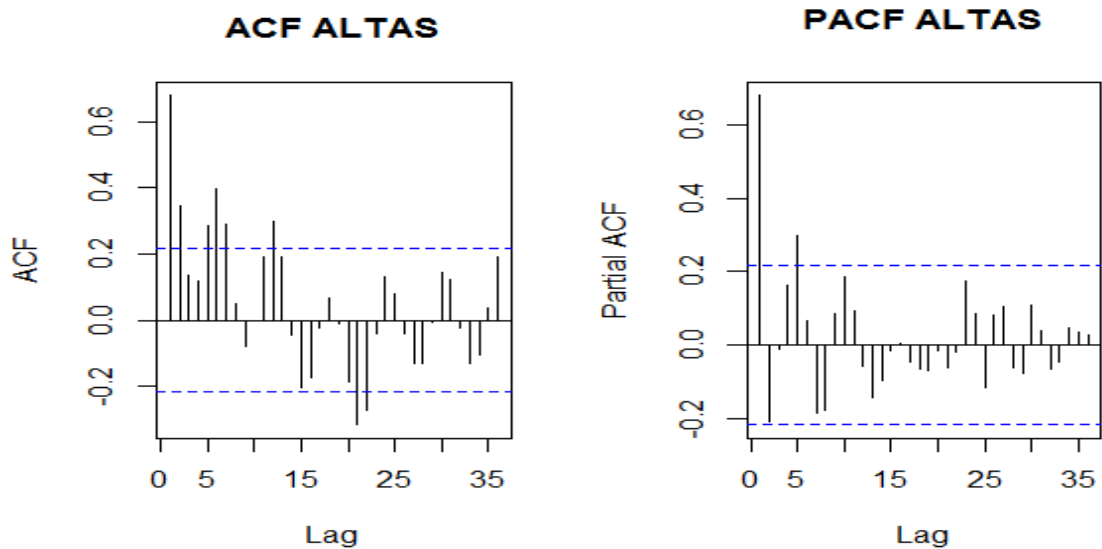


Figura 6: ACP y PACF altas

Haciendo las pruebas de raíz unitaria se toma como referencia el Test Dickey-Füller Aumentado, probando si la serie altas es estacionaria.

$H_0: \rho = 0$ no es estacionaria

$H_1: \rho < 0$ es estacionaria

Variable	Test Dickey-Füller	P-Valor con α 5%	Hipótesis
Altas	1.374	-1.95	H_0: $\rho=0$ no es estacionaria
∇Altas	-1.8877	-1.95	H_0: $\rho=0$ no es estacionaria
$\nabla\nabla^{12}$ Altas	-2.0562	-1.95	H_1: $\rho<0$ es estacionaria

Tabla 3: Test Dickey-Füller aumentado serie original

En la tabla 3 se registran los valores de la prueba Dickey-Füller para determinar las diferenciaciones requeridas hasta que la serie sea estacionaria

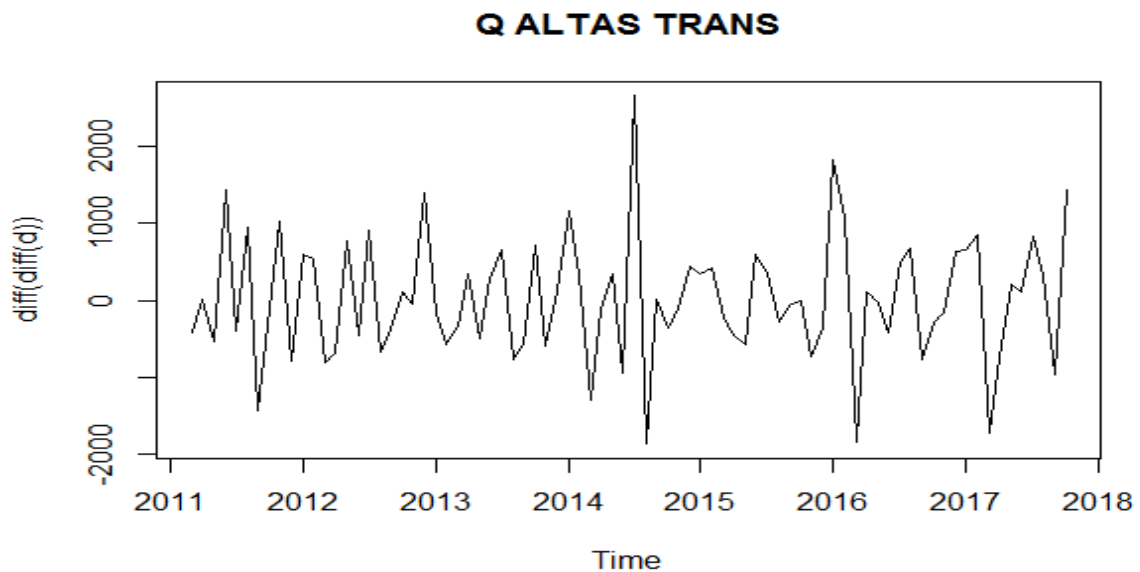


Figura 7: Serie altas transformada

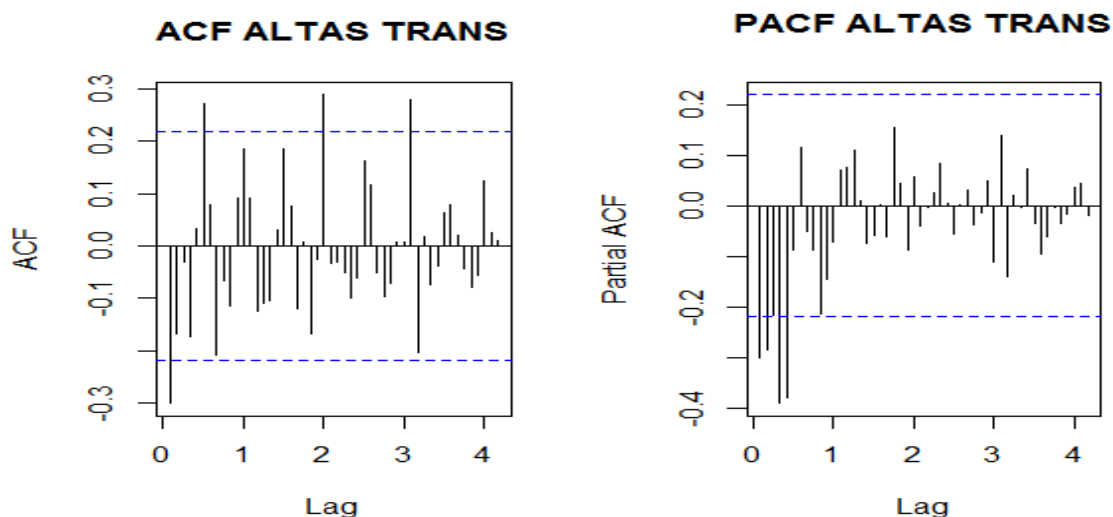


Figura 8: ACF y PACF serie transformada

En la Figura 7 y 8 se observa la serie trasformada con su ACF y PACF, donde se puede determinar posibles órdenes para MA(q) de 1 ó 6 y para AR(p) de 1,2,4,5

Estimación:

Se procede a validar los diferentes modelos SARIMA con la serie diferenciada, de acuerdo al análisis previo de identificación de los gráficos de ACF y PACF, adicional con el apoyo del paquete R, comando autoarima de la librería forecast se estima el mejor modelo que sugiere el programa, a continuación se observan los BICs producto de los modelos estimados.

Modelo	Bic
SARIMA (0,1,0)x(1,1,0) ₁₂ (AO23,AO66)	1059,223
SARIMA (0,1,0)x(1,1,0) ₁₂	1073,419
SARIMA (1,1,1)x(0,1,0) ₁₂	1090,649
SARIMA (1,1,2)x(0,1,0) ₁₂	1094,206
SARIMA (1,1,3)x(0,1,0) ₁₂	1098,165
SARIMA (1,1,5)x(0,1,0) ₁₂	1105,612
SARIMA (6,1,1)x(0,1,0) ₁₂	1108,902
SARIMA (6,1,2)x(0,1,0) ₁₂	1111,191

Tabla 4: Bic de los modelos estimados

Diagnóstico:

Identificando el menor BIC se utilizan las pruebas Box Ljung y Jarque Bera para identificar si existe correlación serial, normalidad.

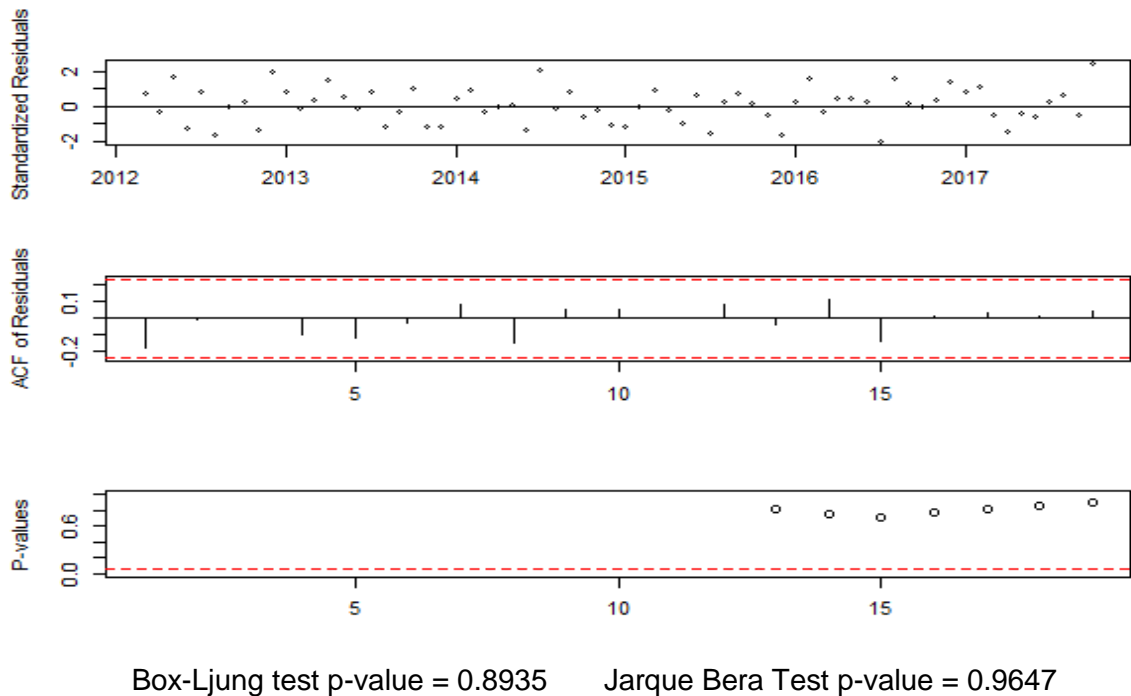


Figura 8: Pruebas Box Ljung y Jarque Bera

El resultado de las pruebas realizadas al modelo SARIMA $(0,1,0) \times (1,1,0)_{12}$, indica que no existe correlación serial con 89% de acuerdo al Ljung Box test y que presenta normalidad al 96% con el test de Jarque Bera, obteniendo Ruido Blanco Gaussiano.

Pronóstico:

Se procede a realizar la predicción de altas del modelo SARIMA $(0,1,0) \times (1,1,0)_{12}$, desde enero del año 2017 hasta octubre del mismo año, donde se contrasta la predicción generada por el modelo frente a la serie real de las altas de banda ancha.

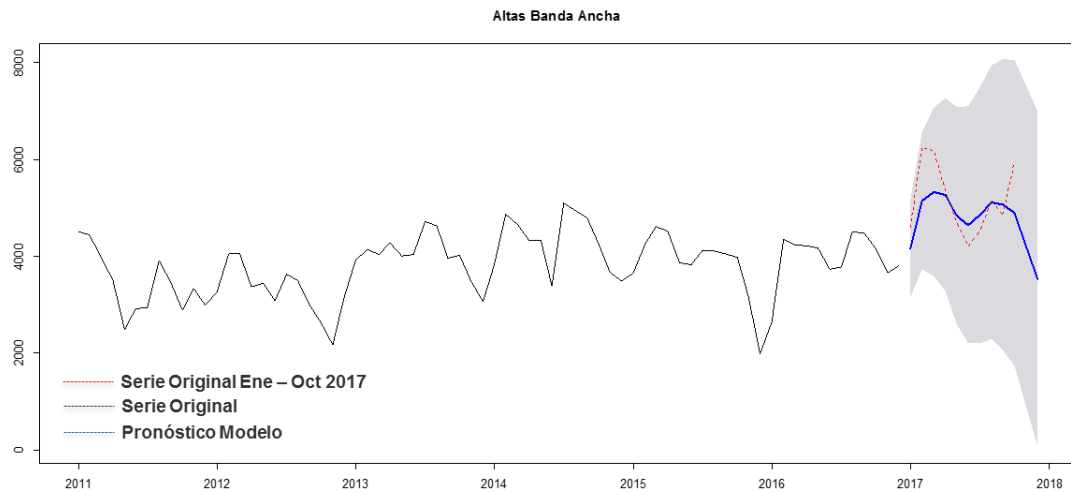


Figura 9: Datos reales vs datos pronosticados

Modelo VAR

Identificación:

Se procede a realizar las gráficas de cada una de las variables originales en el tiempo con su ACF y PACF para realizar un diagnóstico preliminar.

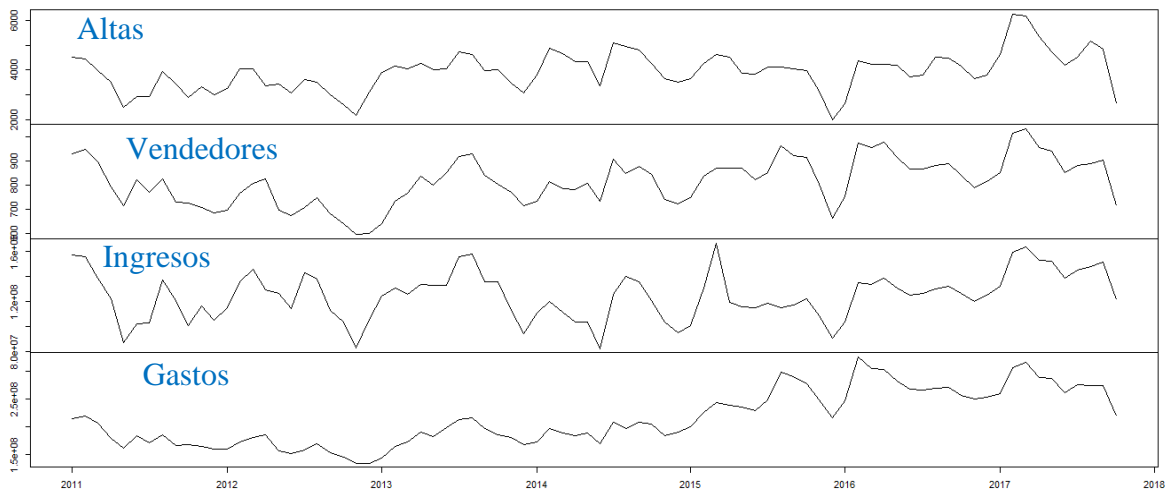


Figura 10: Serie variables seleccionadas

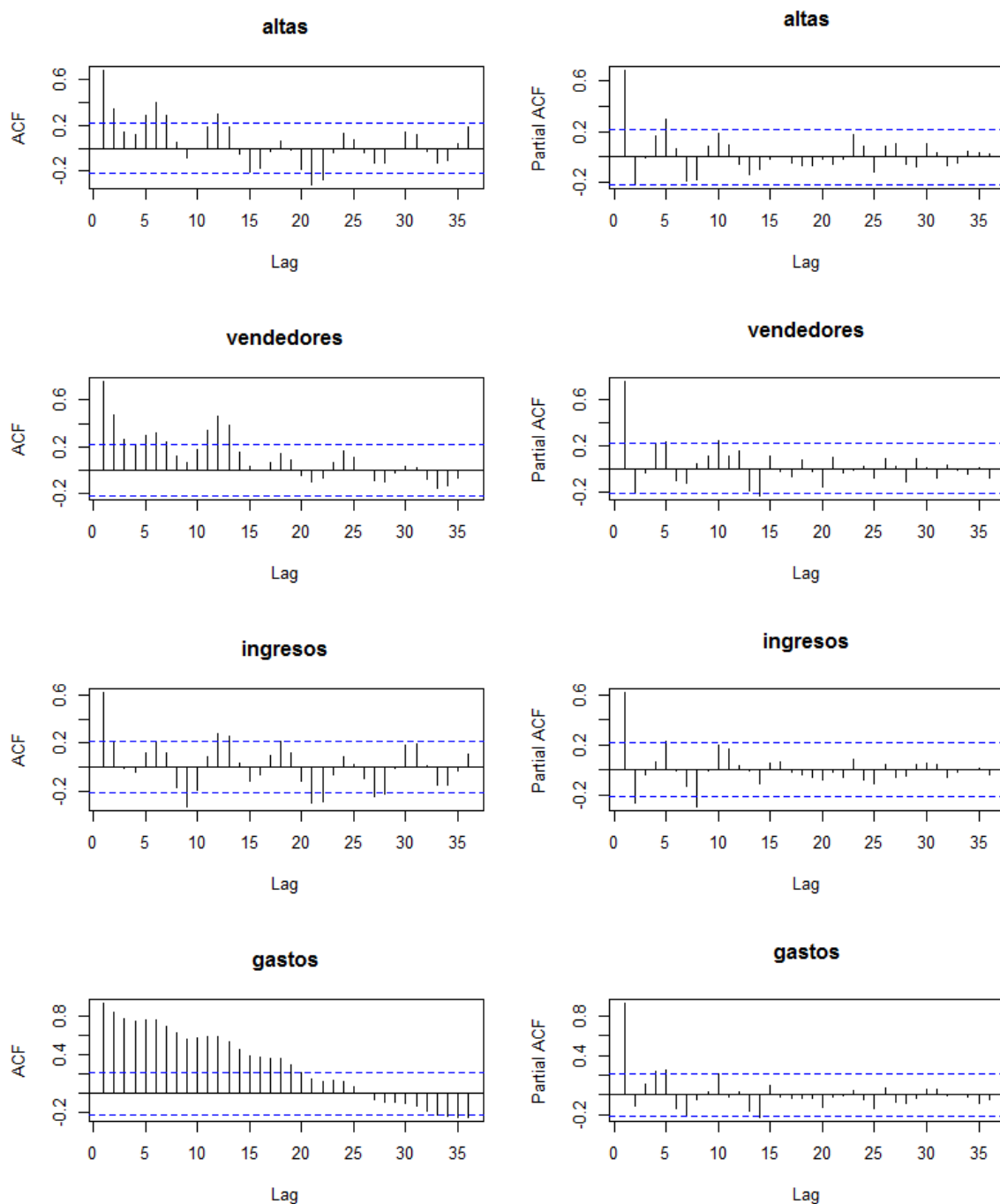


Figura 11: ACF y PACF de variables principales

Validando la autocorrelación simple y parcial de las variables principales, se evidencia un ciclo estacional a 12 meses en la variable **altas**, por lo tanto se procede a utilizar el paquete R con la librería “**VARS**”, para tener en cuenta el ciclo.

Estimación:

Se procede a validar el orden del modelo VAR, utilizando como herramienta el paquete R por medio del comando VARselect de la librería VARS, donde podremos incluir la estacionalidad de la variable altas, a continuación

\$selection					
AIC(n) HQ(n) SC(n) FPE(n)					
10 2 2 2					
\$criteria					
	1	2	3	4	5
AIC(n)	2,80E+07	2,79E+07	2,82E+07	2,85E+07	2,87E+07
HQ(n)	2,91E+07	2,88E+07	2,96E+07	3,01E+07	3,05E+07
SC(n)	3,08E+07	3,02E+07	3,17E+07	3,25E+07	3,33E+07
FPE(n)	1,54E+18	1,32E+18	2,22E+18	3,25E+18	4,87E+18
	6	7	8	9	10
AIC(n)	2,84E+07	2,83E+07	2,75E+07	2,70E+07	2,70E+07
HQ(n)	3,04E+07	3,05E+07	3,00E+07	2,97E+07	2,98E+07
SC(n)	3,35E+07	3,40E+07	3,37E+07	3,37E+07	3,42E+07
FPE(n)	4,51E+18	5,86E+18	4,35E+18	5,20E+18	1,48E+19

Tabla 5: Modelos sugeridos por VARselect

De acuerdo a la Tabla 5 se identifican los modelos que sugiere el paquete R según el HQ, SC, y FPE.

Diagnóstico:

El modelo sugerido por el paquete R comando VARselect es un VAR (2), por lo tanto, se procede a validar los ACF y PACF de los residuales para comprobar si el modelo se ajusta de manera adecuada, para realizar una predicción de los datos, en las figuras 12, 13, 14 y 15 se puede evidenciar que los rezagos de los residuales de las variables no son significativos.

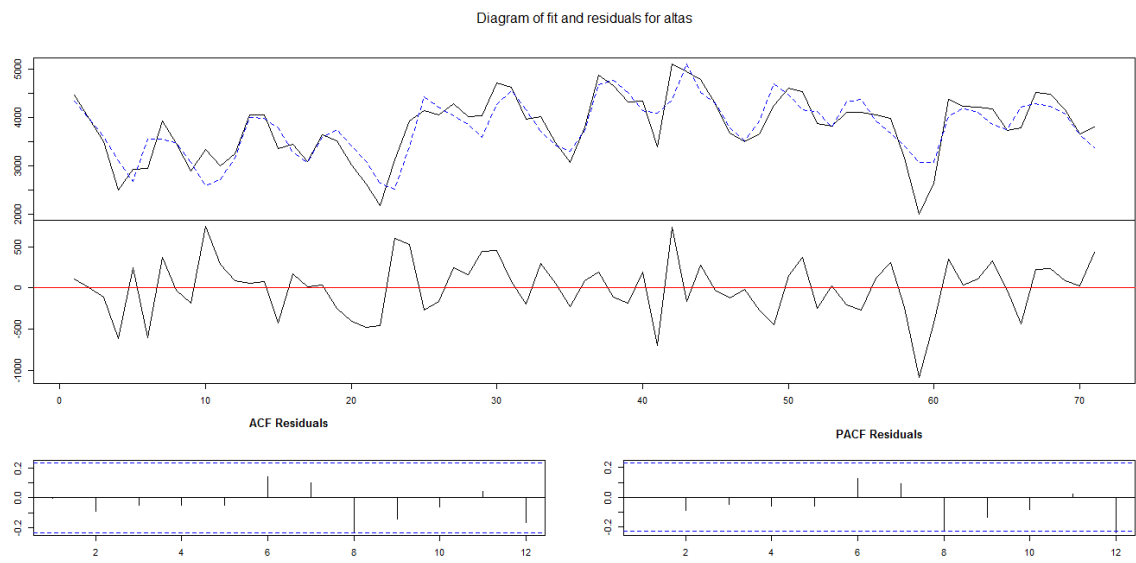


Figura 12: ACF y PACF de los residuos de la variable altas para el modelo VAR(2)

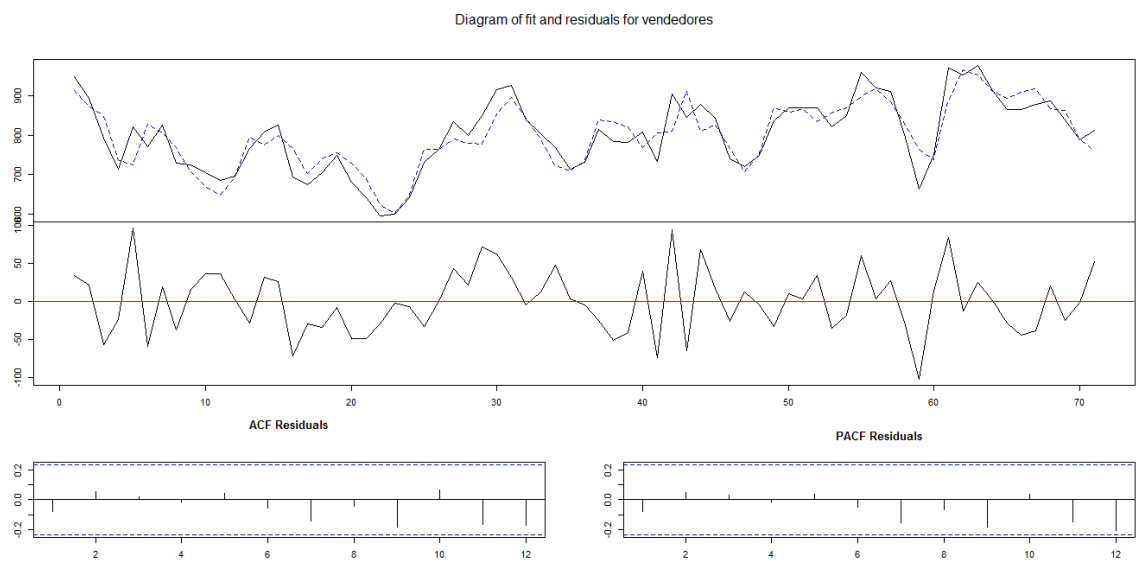


Figura 13: ACF y PACF de los residuos de la variable vendedores para el modelo VAR(2)

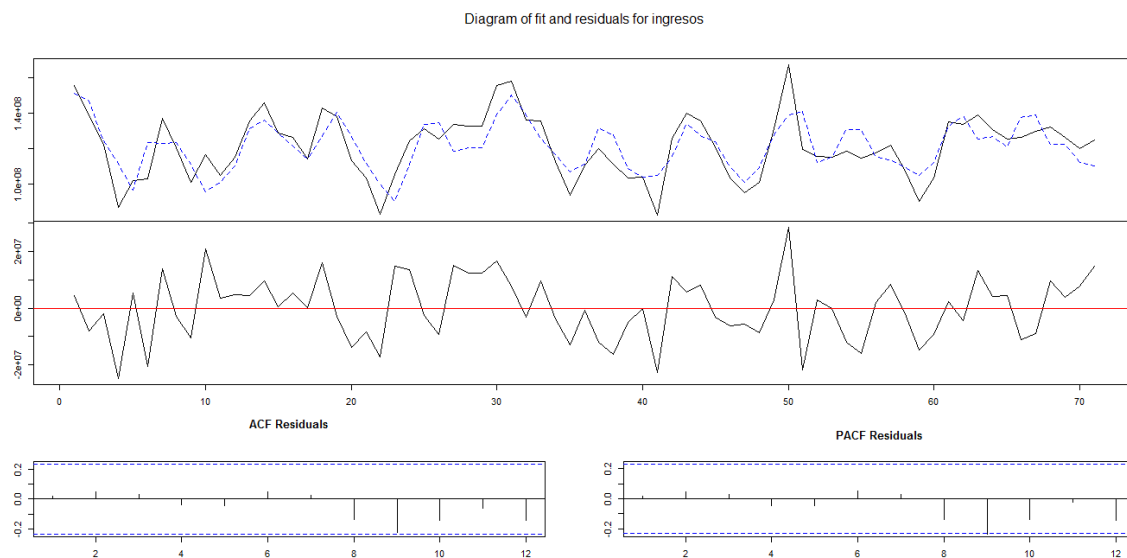


Figura 14: ACF y PACF de los residuales de la variable ingresos para el modelo VAR(2)

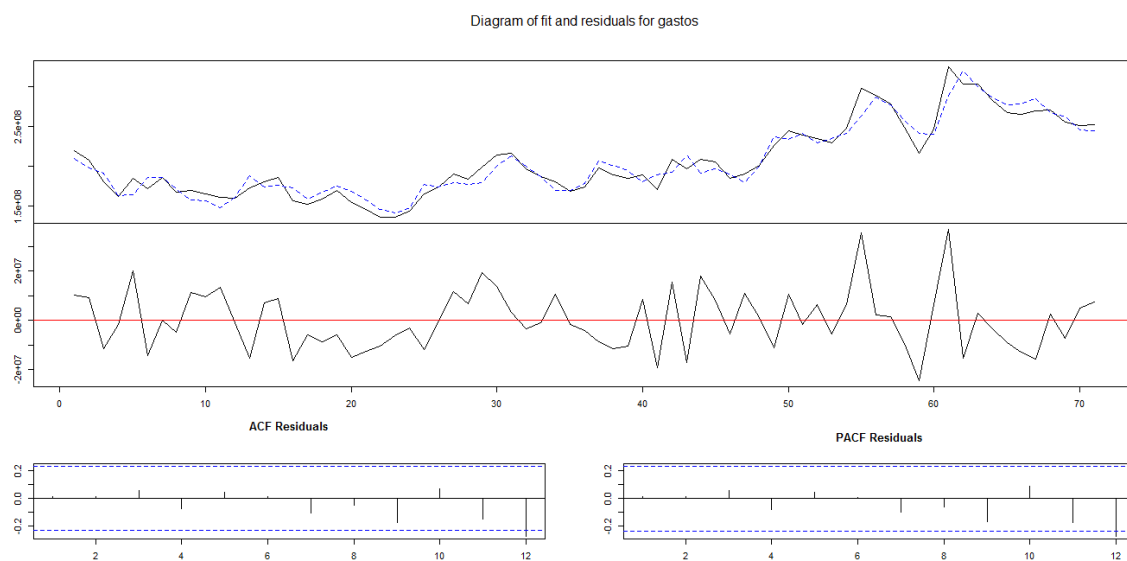


Figura 15: ACF y PACF de los residuales de la variable gastos para el modelo VAR (2)

A continuación, se realizan las pruebas de causalidad donde se valida la influencia pasada y contemporánea de una variable respecto a la predicción de las otras, los resultados de estos test se encuentran en la tabla 6, donde la información pasada de la variable altas no explica las demás variables y la información de las variables vendedores, ingresos y gasto ayudan a explicar las demás variables. De manera contemporánea todas las variables se explican entre sí. Se realizan pruebas de correlación serial en los residuales describiendo los p-valor en la tabla 7.

	VARIABLES	Ho	Value	p-valor
INFORMACION PASADA	altas	La información pasada de altas no ayuda a predecir las variables de vendedores, ingresos y gastos	Se rechaza Ho	0,028
	vendedores	La información pasada de Vendedores no ayuda a predecir las variables de altas, ingresos y gastos	No se rechaza Ho	0,24
	ingresos	La información pasada de ingresos no ayuda a predecir las variables de altas, vendedores y gastos	No se rechaza Ho	0,57
	gastos	La información pasada de gastos no ayuda a predecir las variables de altas, vendedores y ingresos	No se rechaza Ho	0,051
INFORMACION CONTEMPORANEA	altas	a información contemporánea de altas no ayuda a predecir las variables de vendedores, ingresos y gastos	Se rechaza Ho	5.952e-06
	vendedores	La información contemporánea de vendedores no ayuda a predecir las variables de altas, ingresos y gastos	Se rechaza Ho	2.846e-07
	ingresos	La información contemporánea de ingresos no ayuda a predecir las variables de altas, vendedores y gastos	Se rechaza Ho	3.148e-05
	gastos	La información contemporánea de gastos no ayuda a predecir las variables de altas, vendedores e ingresos	Se rechaza Ho	3.957e-07

Tabla 6: Test de causalidad

PRUEBAS DE CORRELACION SERIAL DE LOS RESIDUALES	Ho	Value	p-valor
Breusch-Godfrey LM test	los datos se distribuyen de forma independiente	No se rechaza Ho	0,27
Edgerton-Shukur F test	los datos se distribuyen de forma independiente	No se rechaza Ho	0,9

Tabla 7: Pruebas de correlación serial de los residuales del modelo VAR

Pronóstico:

Se procede a realizar la predicción del modelo VAR (2), de cada una de las variables que acompañan al modelo para el año 2018.

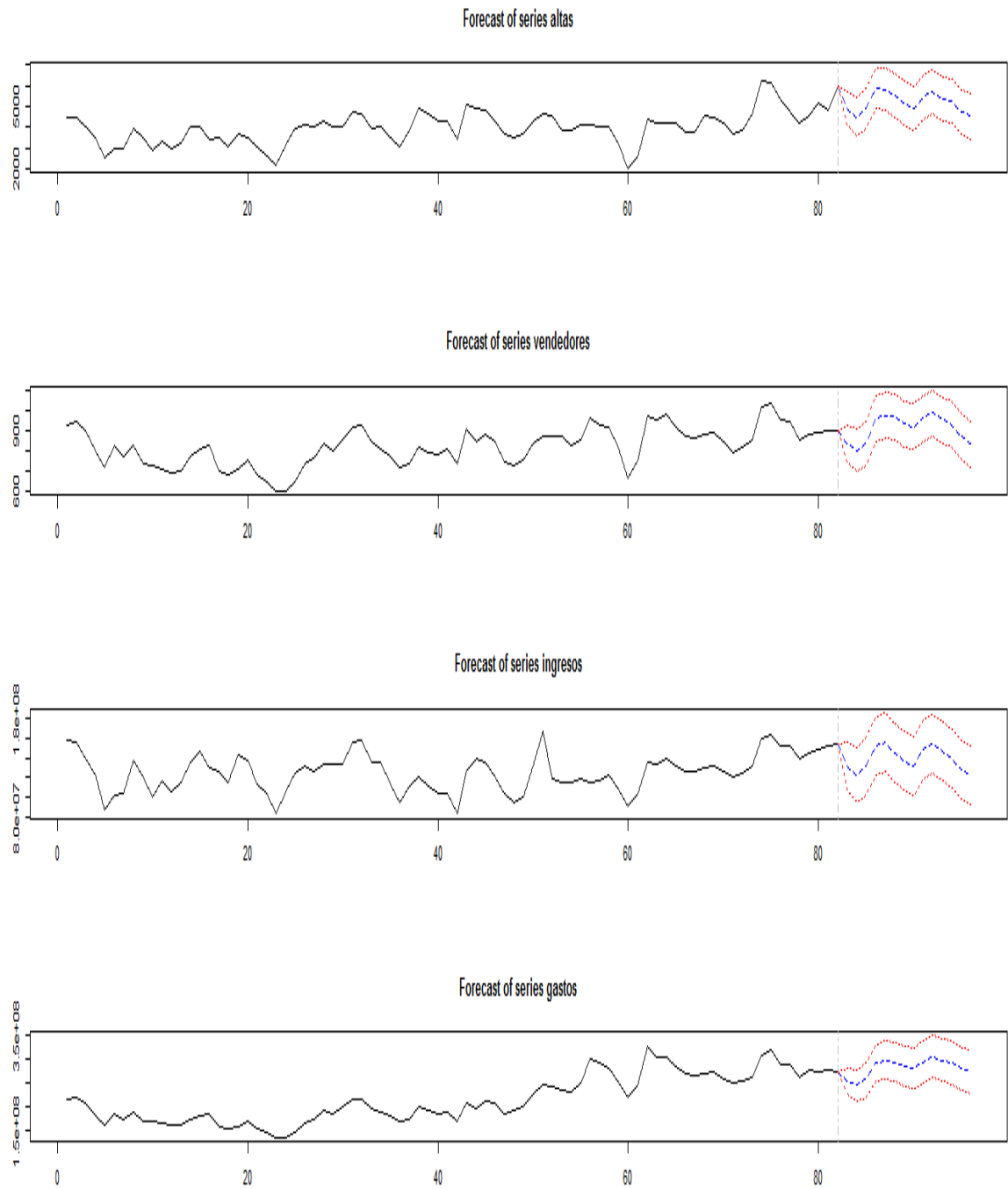


Figura 16: Predicción modelo VAR(2)

5. ANÁLISIS Y RESULTADOS

Con base en la metodología se estima que el mejor modelo univariado para poder predecir el comportamiento de las altas de banda ancha es SARIMA (0,1,0)(1,1,0)₁₂ y para predecir las principales variables que acompañan a las altas el modelo multivariado VAR(2) sería el más apropiado.

Comparando los dos modelos en la figura 17, se evidencia que los datos reales que se dejaron para evaluar los modelos, están más cercanos a los datos de predicción del modelo SARIMA(0,1,0)(1,1,0)₁₂, lo que da a entender que la predicción de las altas se explica mejor respecto a su propio comportamiento.

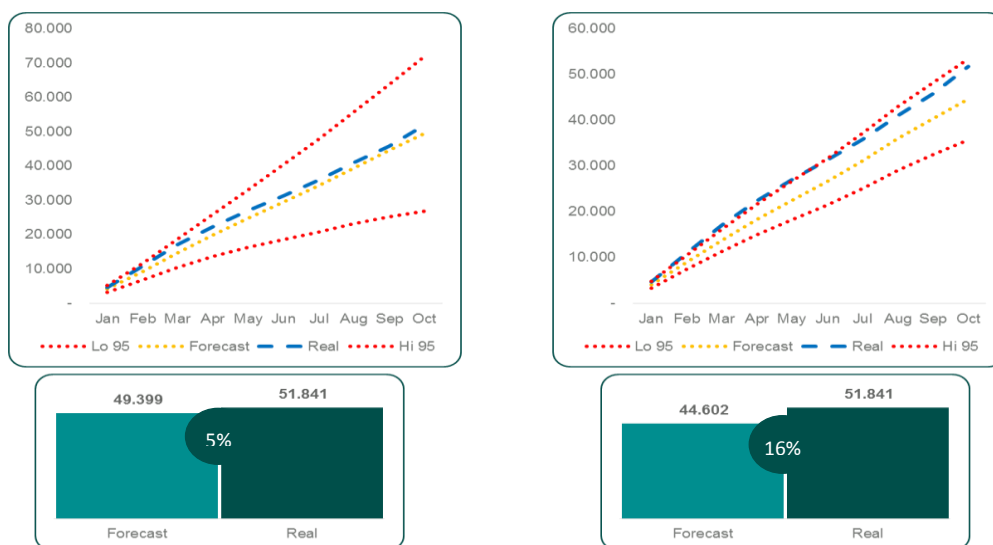


Figura 17: Comparación del modelo SARIMA(010)(110)12 con modelo VAR(2)

Comparando los datos acumulados reales respecto a los predichos, para el modelo SARIMA se tiene una desviación respecto a las altas reales del 5% y para el modelo VAR 16%, confirmando que el modelo para realizar la planeación de la inversión de red con base en el inventario es SARIMA (0,1,0)(1,1,0)₁₂ incluyendo dos AO en los datos (23 y 66) correspondientes a los datos atípicos de los meses noviembre 2012 y junio de 2016. Realizando la predicción se estima que para el año 2018 se tendrán 74,1K altas con un mínimo esperado de 44,1K y un máximo de 104,1K, en la figura 20 se puede observar el pronóstico de las altas.

$$Z_t = e_t + \Phi_{12}e_{t-12} + \theta_0^*A023 + \theta_1^*A066$$

$\{e_t\} \sim \text{RBG}(0; 120578)$
 $Z_t = \nabla \nabla^{12} CP_t$

Ecuación 5: SARIMA(010)(110)12

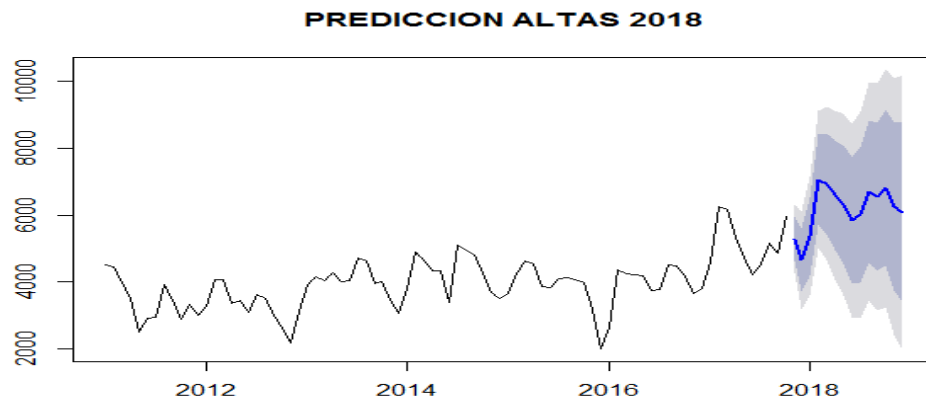


Figura 18: Predicción de altas año 2018

Con las altas estimadas, se realiza el descuento de los 65.000 circuitos de red del inventario de noviembre 2017 obteniendo como mes de inversión junio de 2018 respecto al indicador de vacancia de circuitos, que se estima en un 30%, podemos observar en la figura 19 la estimación de inversión.

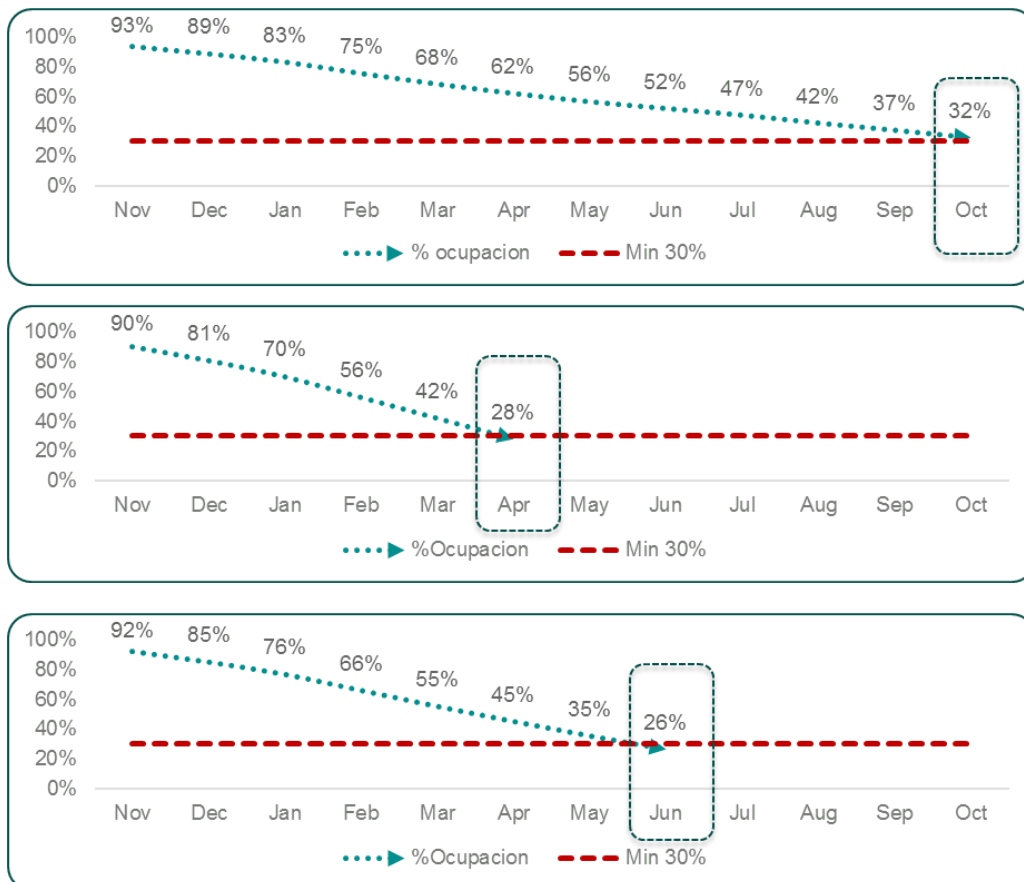


Figura 19: Estimación de inversión

Con base en el modelo VAR(2) y tomando la predicción de las demás variables se puede estimar que para el 2018 se tendrá un promedio de 1032 vendedores, 163 MCOP en ingresos y 329 MCOP de inversión.

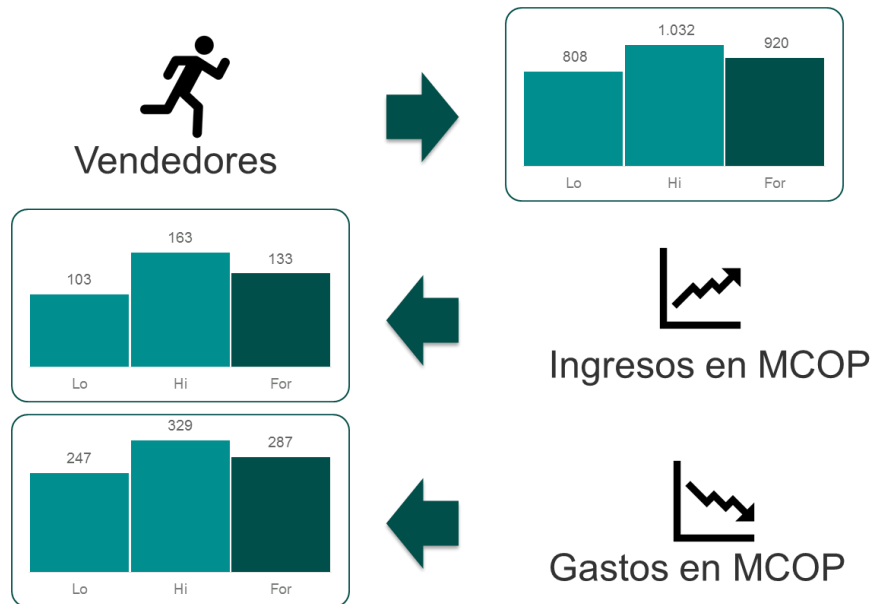


Figura 20: Proyección vendedores, ingresos y gastos

$$Z_t = \hat{\Phi}_0 + \hat{\Phi}_1 Z_{t-1} + \hat{\Phi}_2 Z_{t-2} + e_t$$

Donde $e_t \sim \text{RBG}(0, \Sigma_e)$

$$Z_t = \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{bmatrix}$$

Ecuación 6: VAR2

6. CONCLUSIONES Y RECOMENDACIONES

Con base en los modelos analizados, las variables que evidencian mayor relación respecto a las altas no ayudan a obtener un mejor modelo de predicción, lo que nos sugiere que la predicción de las altas puede realizarse de forma acertada por medio de su historia, utilizando un modelo univariado.

El mejor modelo para poder realizar la predicción de altas es el modelo SARIMA $(0,1,0) \times (1,1,0)$ en donde se pudo contrastar un 5% de desviación de las altas predichas respecto a las altas reales, Descontando las cantidades de altas estimadas al inventario de red y determinado una inversión a mediados del 2018 asumiendo un indicador de vacancia del 30%.

El modelo VAR (2) pese a tener un porcentaje de desviación mayor en la predicción de las altas, nos permite pronosticar de manera simultánea el resto de variables generando un panorama completo e indicando la cantidad de vendedores, ingresos y gastos del 2018 complementando las actividades de comercialización del producto de banda ancha.

De acuerdo con la metodología realizada en este proyecto podemos recomendar que los modelos univariados y multivariados pueden pronosticar de manera eficiente los diversos productos y servicios que comercializan las empresas de telecomunicaciones.

Se sugiere buscar la posibilidad de incluir variables externas a las bases de datos de la empresa, que también están implícitas en el negocio de telecomunicaciones, tales como: La demanda, los hogares y la competencia con el fin de obtener un mejor modelo.

Se recomienda explorar la posibilidad de trabajar las variables Altas, Vendedores, Ingresos y Gastos utilizando un modelo VEC.

7. REFERENCIAS

- Chatfield, C. (1995). *Analysis_of_Time_Series_An_Introduction*. New York: Chapman y hall/crc.
- Gujarati, D., & Porter, D. (2010). *Econometria*. Mexico: McGraw.Hill.
- Mauricio, J. A. (2007). *Introduccion al análisis de series temporales*. Madrid.
- Sarmiento, E. (29 de Abril de 2017). *El espectador*. Obtenido de <https://www.elespectador.com/opinion/el-reves-de-las-privatizaciones-columna-691623>
- Tecnóesfera. (03 de Agosto de 2017). *Colombia registra 28,4 millones de conexiones a banda ancha*. Obtenido de el tiempo: <http://www.eltiempo.com/tecnosfera/novedades-tecnologia/numero-de-conexiones-a-banda-ancha-en-colombia-116374>
- TSAY, R. (2013). *Multivariate Time Series Analysis*. Chicago: Wiley.